



## Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## NOTES AND LITERATURE

### SOME RECENT STUDIES ON VARIATION AND CORRELATION IN AGRICULTURAL PLANTS

FROM the nature of his material the student of agricultural problems has an unexcelled opportunity to collect large masses of statistical data. Domestic animals and plants, particularly the latter, can be easily propagated in vast numbers under conditions controlled in all sorts of ways. Not only the opportunity, but also the desirability, of collecting data on a statistical scale, has been recognized by agricultural investigators from the beginning of experiment-station work in Germany, and still earlier by individual students in this field. Much of the early statistical material relating to agricultural objects or problems still remains unanalyzed and undigested, because of a lack of adequate statistical methods, on the one hand, and a lack of acquaintance on the part of the collector of the data with what mathematical methods did exist for the analysis of such material, on the other hand.

It was obviously to be expected that a system of adequate biometric methods, such as that which has been developed by Professor Karl Pearson, would in due time come to play a conspicuous part in agricultural investigations. This time is coming. One who follows the current literature of agricultural science, in a broad sense of the term, can not fail to be struck with the rapidly increasing use of these mathematico-statistical methods during the last few years. In so far as the methods are correctly and appropriately used this is a most commendable movement. But it must always be kept in mind not to let admiration for the method *per se* blind one as to the real significance and importance of the biological problem attacked. The futility of dealing biometrically with data or problems which lack a sound biological basis is obvious. The indiscriminate application of biometric methods to all kinds of data is easily seen upon critical examination, to have only so much value or validity as resides in the original data themselves. It is particularly important that this point be kept in mind in agricultural work along biometric lines, because of the great ease with which mere statistics can be collected in this field, and the consequent temptation to collect them without critical consideration of their meaning and worth.

It is the purpose of the present review to discuss some of the recent work which has been done along biometric lines with agricultural materials and on problems relating to the science of agriculture. The list of literature at the end of the review based on it does not aim at completeness either in respect to the period or the field covered. Rather it is the aim to indicate the general trend of work in this field and to discuss its points of strength and of weakness.

At the outstart should be mentioned a number of papers which have dealt with the general subject of statistical methods as applied to agricultural material. The general purpose of such papers has been, on the one hand, to call the attention of agricultural workers to the existence of such methods and to the desirability of their use, and, on the other hand, to give some account of the nature of the methods themselves. Here are to be noted the papers of Albrecht, Roemer (introductory portion), Schoute, Quante, Rietz and Smith, and Zaleskigo. The last three papers are especially worthy of attention. The paper of Rietz and Smith gives an excellent elementary discussion of correlation. It further furnishes a most hopeful sign of the rapid development in research standards in agricultural work in this country. Zaleskigo makes keen analytical use of frequency polygons in his breeding work. He calls special attention to the prime importance of not lumping together non-homogeneous material. Rather he urges studying the frequency polygon derived from the progeny of each "pure line" by itself. Then later these separate polygons may, if there is reason for it, be summed together to make a "general population" polygon. But to start with the latter and neglect the biological units (pure lines) which go to make it up is wrong. This insistence on the strict biological or gametic homogeneity of material to be studied by statistical methods is worthy of all commendation.

Quante discusses from a general standpoint some of the problems of variation in agricultural plants. He considers that a definite morphological difference is certainly present between species, varieties or groups when their means differ by five or more times the probable error. He shows that in a number of characters of barley and wheat which he studied the variation is distinctly skew. In a selected strain of rye he found clear evidence of a "normal" or Gaussian symmetrical distribution of variation.

Turning our attention next to special investigations we may

consider different crops separately and take wheat first. Here the studies of Roberts and his students take a leading position. For some years this investigator has been engaged upon a very comprehensive biometrical study of wheat. Only fragments of this work have as yet been published. We may first consider his paper on "A Quantitative Method for the Determination of Hardness in Wheat." An apparatus was devised by which the weight in grams necessary to crush a grain of wheat could be directly determined. The problem was to find out how large a random sample of kernels must be taken in order to reach a reliable result as to mean crushing weight for a variety or strain. Samples of from 100 to 500 kernels each were tested and the mean for each sample determined, two varieties of wheat—a hard and a soft—being used. It is shown that the mean crushing weight diminishes regularly and rapidly as the size of the samples increases, until a minimum at a sample of 450 kernels is reached. Samples of 500 kernels show an increase in mean crushing weight over the 450 kernel sample. Why the *mean* crushing weight should regularly diminish with increasing size of sample is not clear, and is neither explained nor even discussed in the paper. That the *error* of the mean crushing point would diminish with increasing size of sample is obvious. The error of the mean is found, as a matter of fact, to diminish according to a hyperbolic curve. A mathematical discussion of this curve of the errors of the means is given, and examination of the second differential shows that the *rate* of diminution of the error becomes negligible after a sample or group size of 350 kernels. It is then concluded that 350 kernels is a sufficiently large sample to use practically in determining mean crushing points.

Roberts's paper on "Breeding for Type of Kernel in Wheat" is a very thorough and extensive biometrical study of the form of the wheat kernel in many different pure lines or races. Means only are given in this paper, but the amount of measuring and computing involved must have been literally stupendous. Only such a biometrical organization as that maintained at the Kansas Station could have managed it. Data are given on mean length, width, length/width index, volume, weight, and specific gravity of the individual kernel, samples of 500 kernels being taken in 5 separate 100-kernel lots for each pedigree strain (pure line). Also determinations were made of the weight of 100 c.c. of grain, of a packed and a struck bushel of grain, and of the percentage volume not occupied by grain when a 100 c.c. measure is filled

with grain. This last determination was made by the alcohol method. The upshot of this elaborate study is to show that the *shape* of the grains as measured by the length/width index is a very significant factor in determining how wheat will grade according to commercial standards. It is shown that "a difference of at least as high as three pounds per bushel may exist between different pure-bred wheats having identical average kernel-volume and kernel-weight." The final conclusion is that the percentage volume of grain in a packed measure would be a much more just and scientific basis for market grain grading than the present system of test bushel weight. This paper illustrates in a very striking way how the scientific method can solve in a precise and final manner a practical commercial problem.

Lill has made a quantitative study of the relation of size, weight and desirability of kernel to germination in wheat. His data indicate that germination capacity is not correlated with size of kernel, but is correlated with density of kernel. No biometrical analysis of the data is attempted.

Waldron has made an interesting and significant biometrical study of the correlation between weight of grain and other plant characters in oats and wheat, using his own measurements for the former cereal, and published data for the latter. He shows that in oats the mean grain weight per head is *negatively* correlated to a rather high degree with (*a*) number of grains per head, (*b*) length of head and (*c*) length of culm. This obviously leads to a somewhat paradoxical result, namely, that when plump, heavy seed is sown, it is seed which is taken from mother plants which are *below* the average in size and yield. Yet careful experiments, covering a period of years, have shown that planting heavy seed gives increased yields. In other words, a practise which amounts to continued selection of the *poorer* yielding of plants as parents results in *increased* yield in the progeny. This paradoxical result needs analysis by careful pedigree breeding.

Clark has published a general biometrical study on variation and correlation in timothy, the material being gained in connection with the extensive breeding experiments with this grass which have been in progress for some years at Cornell University. The point of chief interest and novelty in the work is that each of the 3,505 plants which furnished the data was under observation during three consecutive years. The material thus gives some basis for an estimation of the relative influence, on the one

hand, of germ-plasm (*i. e.*, germinal determinant factors of whatever sort), which presumably was identical for each plant during the three years, and environmental factors, on the other hand, in determining observed degrees and kinds of variation in the adult organism. The results taken as a whole show that what might be called the general variation *facies* of a population of *Phleum* must depend to a very high degree upon "nurture" rather than "nature." The degree of variation, the degree of skewness of the variation curves, the closeness of correlation between different characters of the plant—all these are changed by general environmental conditions to a marked extent. Thus to take an example: the coefficient of correlation between weight and height of plant is given as  $.274 \pm .011$  in 1905 and as .718 in 1907. This is a *relative* change of nearly 200 per cent. In another case a significant positive correlation one year becomes significantly negative two years later. In general the heights (or weights) of timothy plants in any one year are correlated with the heights (or weights) of the same identical plants in another year only to about the degree indicated by a coefficient of around .5, which is but 50 per cent. of perfect correlation.<sup>1</sup> In other words, it appears on the basis of this result that in determining what a given timothy plant shall be like next year in respect to such characters as height and weight the innate constitutional, hereditary factors within the plant are on the whole of neither greater nor less importance than external environmental circumstances. In this case, and in respect to the characters dealt with, "nature" and "nurture" are about evenly balanced, with what advantage there is on the side of "nurture." The author emphasizes the practical significance of a result of this kind to the man who is carrying on selective breeding, and who obviously must make his selections at the outstart on the basis of the visible somatic characters as they are developed at the particular place and time at which he is doing his selecting. The paper is unfortunately marred by arithmetic errors.

It is a well-known fact that European workers (other than English), generally speaking, have very little acquaintance with biometric technique. A good example of this fact is afforded by a paper of Grabner on the problem of correlated variation in barley. The investigator desired to learn what relation existed between the economically valuable characters of this cereal. He collected a vast lot of statistical data regarding such characters

<sup>1</sup> Cf. Clark's Table VIII.

as yield of grain, hectoliter weight, weight of 1,000 kernels, size of kernel, protein content and "mealiness" or softness of grain. Instead of proceeding by the straightforward method of forming a correlation table and deducing therefrom the coefficient of correlation the author follows the laborious, inaccurate and inconclusive plan of averages. Virtually what is done is to calculate the observed regression line of one character on another. The general result reached, though in no wise critically supported by the evidence presented, is that all of the purely physical characters are correlated together to a high degree. The chemical and chemico-physical characters protein content and "mealiness" are not demonstrably (by the method used) correlated with other characters, though they are mutually definitely correlated. The chief scientific value of the paper is to illustrate in a striking manner how crude and clumsy were pre-Galtonian methods of attacking a simple statistical problem.

Turning now to corn, we have a number of studies of a more or less biometrical character. Apart from the primarily genetic studies on maize of East, Shull, Collins, and Pearl and Surface which are quantitative in character and to some extent<sup>2</sup> biometric in the treatment of the data, there have appeared recently two special studies on variation and correlation in this plant. The first of these is the paper of Rietz and Smith and the second that of Ewing. The objects of the two papers are apparently somewhat dissimilar. Ewing's is primarily a biological investigation, whereas Rietz and Smith apparently desire primarily to set forth the method of measuring correlation, and incidentally to illustrate these principles by means of some corn data which they have on hand. The only general result of particular biological significance brought out in the work of Rietz and Smith is that the degree of correlation between various ear characters (length, circumference, number of rows, weight) is very markedly influenced by environmental conditions surrounding the growing crop. This paper is to be commended for its clear exposition of the method of calculating a correlation coefficient.

Ewing's paper contains more matter of general biological interest. Especially to be mentioned is the valuable discussion of the literature of correlation. The general problem which formed the basis of this investigation was to learn in how far the

<sup>2</sup>Shull gives some very interesting data in the form of variation constants (mean, standard deviation, and coefficient of variation) for variation in number of rows on ear in pure and cross-bred ( $F_1$  and  $F_2$ ) maize.

determination of statistical correlations between different parts of the maize plant might be of use to the practical breeder. The general conclusion to which the author comes in regard to this point is as follows:

Considerable study of the subject has forced upon the writer the belief that it is improbable that much use can be made of correlation in practical breeding. There are rare cases in which the coupling of unit characters may aid the breeder in making selections at an early period, but the existence of correlation in the fluctuating variability of two different characters is not likely to prove of much assistance. Nothing more than a moderate degree of correlation is likely to be found in these cases, unless some such relation as cause and effect exists between them. This is especially true of correlation between seed production and other characters, since the former depends upon a large number of other characters and conditions.

The correlations studied were those of weight of grain per plant (measuring *yield*) with each of the following characters: (1) Diameter of stalk, (2) length of leaf, (3) breadth of leaf, (4) height of mature plant, (5) height of seedling, (6) number of internodes, (7) average length of internodes, (8) percentage of internodes below the ear, (9) length of ear at appearance of silks, (10) date of appearance of tassel, (11) date of appearance of pollen, (12) date of appearance of silks, (13) duration of flowering period (pistillate flowers) in days, (14) number of branches in the tassel.

The coefficient for correlations 1-6, inclusive, 9, 10 and 12, are, in each case, from 5 to 19 times the respective probable errors. They are thus statistically significant. In view of this fact the statement in the general discussion of results that "in most cases the coefficient of correlation is so small that it is probably not worth while to try to classify it or even to conclude that there is correlation," seems not to have been very well considered. The same criticism is to be made against the paper of Clark discussed above. These authors appear to overlook the fact that whether a correlation is statistically significant (*i. e.*, whether correlation "*exists*") depends not upon its absolute value, but upon its relation to its probable error. A coefficient of  $.0009 \pm .0001$  would be to a high degree of probability *statistically* significant, though absolutely small.

The garden pea (*Pisum sativum*) has been the subject of several recent biometric studies. At the Massachusetts Station Waugh and Shaw have been for some time engaged in an

investigation of inheritance in this form, conducted along biometric lines. In their first paper here reviewed they present variation data regarding the four following characters: length of vine, number of pods per vine, length of pod, number of peas per pod, and total peas. The raw data are not given and the discussion is very meager. Graphs of the variation curves are given, but instead of making these plottings of the actual frequency data as polygons, the authors connect the plotted points by free-hand sweeping curves. This is certainly a simple and expeditious, if somewhat naïve, method of curve-fitting! It is much to be regretted that such an inadequate, and indeed absolutely incorrect, method of presentation of statistical results should have been resorted to. In the discussion of heredity stress is laid upon the varying degrees of prepotency observed in the transmission of characters by individual plants. To measure this a new "coefficient of heredity" is proposed. The formula for this is

$$C = 1/\sigma D,$$

where  $C$  is the proposed coefficient,  $\sigma$  the standard deviation of offspring and  $D$  the difference between the parental character and offspring mean of the same character. It is obvious that the more nearly the offspring are like each other, and like the parent the larger will  $C$  become. It is somewhat unfortunate that this is called a "coefficient of heredity," since this term is in common biometrical usage for a very different constant. Indeed, in their own paper Waugh and Shaw use this term not only for their proposed constant, but also for the correlation coefficient between parent and offspring. A satisfactory measure of *individual* (not *average*) prepotency is a thing which is badly needed in breeding work. While the constant  $C$  proposed by Waugh and Shaw meets some of the conditions which such a measure must fulfill, it unfortunately appears to be of rather restricted significance and usefulness. The numerical value which it takes for different characters are not comparable one with another. The reason, obviously, is because the numerical value will change in accordance with the absolute rather than the relative variability of the character. An elephant and a mouse each equally prepotent with reference to the transmission of any character, say skull breadth, would have very different values of  $C$  for this character. Further, the constant becomes rather difficult to manage in cases of biparental inheritance, or

in those cases of undoubted prepotency, which are of the greatest interest and importance both theoretically and practically, wherein the prepotent individual does not itself have the character with regard to which it is prepotent expressed in its own soma. An example here is the dairy bull, prepotent in respect to milking qualities.

A continuation of this work on peas is reported in the second paper by the same authors. Data are presented showing the relation between observed variability and environmental (seasonal) conditions. The interesting point is brought out that there is less variation, and a higher correlation between parent and offspring, in respect to vine length, than in respect to either number of pods per vine or total peas per vine.

Roemer gives a very detailed biometrical study of pure lines in peas. The work is essentially a confirmation, with another plant, of Johannsen's epoch-making investigations on beans, though it lacks any extensive studies on the effect of selection within the pure line. The essential objective point of Roemer's research is rather to determine the biometric characteristics of pure lines as such in relation to the general population. Among the more important general results are the following:

1. The different biotypes in a population arrange themselves in frequency distributions in accord with Quetelet's law.

2. No relation was found to exist between the variability of the biotypes (*i. e.*, variation within the general population) and variation within the pure lines.

Shaw has made a very thorough biometric study of variation in the Ben Davis variety of apples and presents a mass of data of considerable general biological interest. When one recalls that commercial apple varieties are propagated by vegetative processes entirely, the importance of a careful study of this variation under different environmental conditions is obvious. Shaw shows that the mean size and shape of apples of the Ben Davis variety are distinctly different for different trees of the same orchard and even for the different parts of the same tree. There are very marked differences in apples of this variety in respect to size and shape characters when they are grown under widely different soil and climatic conditions. In the south the Ben Davis is a short round apple; in the north it is an elongated apple. Not only are the means different in different environments, but also the variability (as measured by the coefficient of variation) is changed. This paper of Shaw's, while itself purely descrip-

tive, is of great value not only for the interesting data regarding variation which it presents, but also in indicating clearly the rich reward which may be expected to follow a combined experimental and biometric attack upon the fundamental biological problem of the effect of stock on scion.

In the papers so far discussed there has been in every case some attempt at biometric analysis of the raw statistical data. There are constantly appearing in agricultural literature papers in which a great mass of first-class statistical material on variation and correlation in agricultural plants is presented but not analyzed biometrically, or only incompletely so. Examples of this are found (to mention but two) in the interesting papers of Kohler on potatoes and Westgate on alfalfa. A conspicuous instance of failure to make profitable use of elementary biometrical methods is seen in the paper of Stockberger and Thompson on hops. These authors put their data in form for calculating variation and correlation constants (*e. g.*, they give a correlation table for the correlation between number of vines to the hill and yield per hill) but do not determine the constants.

It is evident from what has preceded that biometrical methods are rapidly gaining a place among the agricultural investigator's working tools. Keeping always in mind the caution expressed at the beginning of this article that biometric zeal be not allowed to outrun biological discretion this movement merits only commendation and further encouragement. The agricultural investigator has an almost unique opportunity to make significant and profitable application of biometric methods of research.

RAYMOND PEARL.

#### REFERENCES TO LITERATURE

1. Albrecht, K. Die Fehlerwahrscheinlichkeitsrechnung und ihre Anwendung auf die Pflanzenzüchtung. *Fühlings Landw. Ztg.*, 1908, pp. 577-585.
2. Clark, C. F. Variation and Correlation in Timothy. *Cornell Agr. Expt. Stat.*, Bul. 279, pp. 301-350, 1910.
3. Collins, G. N. The Value of First Generation Hybrids in Corn. *U. S. Dept. Agr., Bur. Plant Ind.*, Bul. 191, pp. 7-45, 1910.
4. East, E. M. A Mendelian Interpretation of Variation which is Apparently Continuous. *AMER. NAT.*, Vol. XLIV, pp. 65-82, 1910.
5. Ewing, E. C. Correlation of Characters in Corn. *Cornell Univ. Agr. Expt. Stat.*, Bul. 287, pp. 67-100, 1910.
6. Grabner, E. Wechselbeziehungen zwischen den wertbestimmenden Eigenschaften der Braugerste. *Jour. f. Landw.*, Bd. 57, pp. 321-350, 1909.

7. Kohler, A. R. Potato Experiments and Studies at University Farm in 1909. Minn. Agri. Expt. Stat., Bul. 118, pp. 65-141, 1910.
8. Lill, T. G. The Relation of Size, Weight and Density of Kernel to Germination in Wheat. Kansas Agr. Expt. Stat., Circ. 11, pp. 1-8, 1910.
9. Pearl, R., and Surface, F. M. Experiments in Breeding Sweet Corn. Me. Agr. Expt. Stat., Ann. Rpt., 1910, pp. 249-307.
10. Quante. Variationsstatistische Untersuchungen über den Bau der Getreidearten unter Zergründelung der Kollektivmasslehre. Die landw.; Versuchs-Stationen, 1910, pp. 121-162.
11. Rietz, H. L., and Smith, L. H. On the Measurement of Correlation with Special Reference to some Characters of Indian Corn. Ill. Agr. Expt. Stat., Bul. 148, pp. 291-316, 1910.
12. Roberts, H. F. A Quantitative Method for the Determination of Hardness in Wheat. Kansas Agr. Expt. Stat., Bul. 167, pp. 371-390, 1910.
13. Roberts, H. F. Breeding for Type of Kernel in Wheat. Kansas Agr. Expt. Stat., Bul. 170, pp. 99-138, 1910.
14. Roemer, Th. Variabilitätsstudien. *Arch. Rass.- u. Gesellsch. Biol.*, 7 Jahrg., pp. 397-469, 1910.
15. Schoute, J. C. Die Fehlerwahrscheinlichkeitstheorie für die Praxis der Versuchsstationen. *Die landw. Vers.-Stat.*, Bd. LXX, pp. 161-180, 1909.
16. Schoute, J. C. Zur quantitativen Reinheitsbestimmung von Leinkuchen und Leinkuchennmehl. *Ibid.*, pp. 171-247, 1909.
17. Shaw, J. K. Variation in Apples. Twenty-second Ann. Rpt. Mass. Agr. Expt. Stat., Part I, pp. 194-213, 1910.
18. Shull, G. H. Hybridization Methods in Corn Breeding. *Amer. Breeder's Mag.*, Vol. I, pp. 98-107, 1910.
19. Stockberger, W. W., and Thompson, J. Some Conditions Influencing the Yield of Hops. U. S. Dept. Agr., Bur. Plant Ind., Circ. 56, pp. 1-12, 1910.
20. Waldron, L. R. A Suggestion Regarding Heavy and Light Seed Grain. *AMER. NAT.*, Vol. 44, pp. 48-56, 1910.
21. Waugh, F. S., and Shaw, T. K. Variation in Peas. Twenty-first Ann. Rpt. Mass. Agr. Expt. Stat., Part II, pp. 167-173, 1909.
22. Waugh, F. A., and Shaw, J. K. Plant Breeding Studies in Peas. Twenty-second Ann. Rpt. Mass. Agr. Expt. Stat., Part I, pp. 168-175, 1910.
23. Westgate, J. M. Variegated Alfalfa. U. S. Dept. Agr., Bur. Plant Ind., Bul. 169, pp. 1-63, Pl. I-IX, 1910.
24. Zaleskiego, E. I. Zastosowanie Wielokatow Czystotliwosci do selekcyi roslin. (Application of Frequency of Polygons in Plant Breeding.) Krakowie, 1909, pp. 1-32. Autoreferat in German, pp. 1-6.

## ON SEX-CHROMOSOMES IN HERMAPHRODITISM

RESEARCHES by Boveri and his pupils have shown in certain nematodes, as in arthropods, the existence of two sorts of spermatozoa, one of which contains one more chromosome or chromosome